Estimating distances via connectivity in wireless sensor networks

Baoqi Huang[1], Changbin Yu[2,3]*, Brian D. O. Anderson[2] and Guoqiang Mao[4]

[1] Inner Mongolia University, Hohhot, China
[2] Australian National University and NICTA Ltd, Canberra, ACT, Australia
[3] Shandong Computer Science Center, Jinan, China
[4] University of Sydney and NICTA Ltd, Sydney, NSW, Australia

ABSTRACT

Distance estimation is vital for localization and many other applications in wireless sensor networks. In this paper, we develop a method that employs a maximum-likelihood estimator to estimate distances between a pair of neighboring nodes in a static wireless sensor network using their local connectivity information, namely the numbers of their common and non-common one-hop neighbors. We present the distance estimation method under a generic channel model, including the unit disk (communication) model and the more realistic log-normal (shadowing) model as special cases. Under the log-normal model, we investigate the impact of the log-normal model uncertainty; we numerically evaluate the bias and standard deviation associated with our method, which show that for long distances our method outperforms the method based on received signal strength; and we provide a Cramér–Rao lower bound analysis for the problem of estimating distances via connectivity and derive helpful guidelines for implementing our method. Finally, on implementing the proposed method on the basis of measurement data from a realistic environment and applying it in connectivity-based sensor localization, the advantages of the proposed method are confirmed.

KEYWORDS

distance estimation; maximum-likelihood estimator; Cramér-rao lower bound; generic channel model; received signal strength

***Correspondence**

Changbin Yu, RSISE Building 115, Australian National University, Canberra ACT 0200, Australia.
E-mail: Brad.Yu@anu.edu.au

1. INTRODUCTION

Wireless sensor networks, composed of hundreds or thousands of small and inexpensive nodes with constrained computing power, limited memory, and short battery lifetime, can be used to monitor and collect data in a region of interest. Accurate and low-cost node localization is important for various applications in wireless sensor networks, and thus, great efforts have been devoted to developing localization algorithms, including distance-based algorithms and connectivity-based algorithms [1]. Distance-based localization algorithms rely on distance estimates and can achieve relatively good localization accuracy; if distance estimates are unavailable or suffer from huge errors, connectivity-based localization algorithms are applied but generally achieve coarse-grained localization accuracy. Besides, distance estimation is vital for sensor network management, such as topology control [2,3] and boundary detection [4].

In reality, distance estimation can be realized by using information such as received signal strength (RSS), time of arrival (TOA) and time difference of arrival (TDOA) [1]. The RSS method (using RSS measurements) depends on low-cost hardware and only provides coarse-grained distance estimates; by contrast, the TOA and TDOA methods can provide distance estimates with higher accuracy at the cost of extra hardware, but because of cost constraints, it is impractical to equip all sensors in a large-scale sensor network with extra hardware to obtain accurate distance estimates and thus accurate location estimates. Further, although a number of connectivity-based localization algorithms have been proposed, for example, see [5–8], achieving high localization accuracy usually demands a comparatively large number of anchor nodes, hereafter termed simply *anchors*, whose positions are known *a priori* (accordingly, we term other nodes whose positions are not known and need to be determined as *sensors*). In this paper, we shall propose an

attractive distance estimation method that does not rely on extra hardware but provides comparatively accurate distance estimates.

In a static wireless sensor network, two nodes are termed one-hop or immediate neighbors as long as they can directly communicate with each other. An intuitive observation shows that with a higher probability, two geographically close nodes share more common immediate neighbors than two distant nodes. We quantify and exploit this observation to develop a maximum-likelihood estimator (MLE) for estimating the distance between any pair of neighboring nodes on the basis of their local connectivity information. Herein, local connectivity information refers to the numbers of common and non-common immediate neighbors associated with a pair of neighboring nodes. Because only elementary computations and local connectivity information are involved on each node, the proposed method is energy efficient and totally distributed.

In this paper, we present the distance estimation method under a generic channel model, including the ideal unit disk (communication) model and the more realistic log-normal (shadowing) model as special cases (see Section 2 for further details). Then, we take the log-normal model for example to demonstrate the proposed method: the impact of uncertainties in the log-normal model is examined; the bias and standard deviation of derived distance estimates are numerically evaluated; the proposed method, although not comparable with such fine-grained distance estimation techniques as TOA and TDOA, outperforms the well-known RSS method for long distances; the influences of various factors on the problem of estimating distances via connectivity are analyzed on the basis of the Cramér–Rao lower bound (CRLB), and useful guidelines for implementing the proposed method in reality are also derived; and finally, on implementing the proposed method on the basis of measurement data in a realistic environment and also applying it in connectivity-based sensor localization, the advantages of the proposed method are confirmed.

Prior to our work, [9,10] came up with the methods of estimating distances on the basis of the same idea as ours. The neighborhood intersection distance estimation scheme (NIDES) presented in [9] heuristically relates the distance, for example, from node A to node B, to an easily observed ratio, that is, the number of their common immediate neighbors to the number of immediate neighbors of A, and then performs the distance estimation at node A using this ratio and other *a priori* known information. The NIDES assumes the unit disk model and uniformly and randomly deployed wireless sensor networks. Its enhanced version presented in [10] adapted the ratio by taking into account the number of immediate neighbors of node B and heuristically stated that the NIDES could be applied in arbitrary radio models. Although it turns out that the enhanced NIDES leads to the same solution as ours, their entire treatment rests on empirical observations and heuristic formulations rather than theoretical foundations. In comparison with their work, the *contributions* of this paper are as follows: (i) a statistical model is formally established for the distance estimation problem, and an MLE solution with mathematical proofs of the correctness is provided; (ii) the problem is considered under a generic channel model widely used in the literature, including the more realistic log-normal model; (iii) the performance of the proposed method is comprehensively analyzed under the log-normal model in terms of model uncertainties, bias, standard deviation and root mean square error (RMSE); (iv) a CRLB analysis is carried out for the problem of estimating distances via connectivity; and (v) it is shown that the proposed method contributes to the quality of connectivity-based sensor localization.

The remainder of the paper is organized as follows. Section 2 introduces the network model and the (radio) channel model. Section 3 proposes the method under a generic channel model. Under the log-normal model, Section 4 analyzes the performance of the proposed method; Section 5 provides a CRLB analysis for the general distance estimation problem using connectivity; Section 6 implements the proposed method using the measurement data from a real environment; Section 7 reports the contributions of the proposed method to connectivity-based sensor localization by simulations. Finally, we conclude the paper in Section 8.

2. SYSTEM MODEL DESCRIPTION

This section briefly introduces the system model we shall use, including the network model and the channel model. Throughout this paper, we shall use the following mathematical notations: $\Pr\{\cdot\}$ denotes the probability of an event, and $E(\cdot)$ denotes the statistical expectation.

2.1. Network model

In static wireless sensor networks, nodes are often assumed to be randomly and uniformly distributed on account of the random nature of the network deployment. A homogeneous Poisson process provides an accurate model for a uniform distribution of nodes as the network size approaches infinity. Therefore, we consider *a static wireless sensor network that is deployed over an infinite plane according to a homogeneous Poisson process of intensity* λ.

2.2. Channel model

Let P_{T} be the transmitted signal power by a transmitter and $P_{\mathrm{R}}(d)$ be the received signal power by a receiver located at distance d from the transmitter. According to [11], the *log-normal model* predicts $P_{\mathrm{R}}(d)$ to be log-normally distributed and is typically modeled as follows:

$$P_{\mathrm{R}}(d)(\mathrm{dBm}) = \overline{P_{\mathrm{R}}}(d_0)(\mathrm{dBm}) - 10\alpha \log_{10} \frac{d}{d_0} + Z \quad (1)$$

$$\overline{P_{\mathrm{R}}}(d) = \frac{\tau^2 G_{\mathrm{T}} G_{\mathrm{R}} P_{\mathrm{T}}}{(4\pi)^2 d^{\alpha}} \tag{2}$$

where $\overline{P_{\mathrm{R}}}(d_0)$ (dBm) is the mean received signal power in dB m at a reference distance d_0, α is the path loss exponent, G_{T} and G_{R} are the transmitter and receiver antenna gains, τ is the wavelength of the propagating signal in meters, and Z is a random variable representing the shadowing effect, normally distributed with mean zero and variance σ^2_{dB}. Typically, α can vary between 2 in free space and 6 in heavily built urban areas, and σ_{dB} is as low as 4 and as high as 12 according to [11].

If $P_{\mathrm{R}}(d)$ is above some specified value P_{c}, the receiver is able to communicate with the transmitter. Particularly, if $\sigma_{\mathrm{dB}} = 0$, the log-normal model is equivalent to the unit disk model with the transmission range

$$r = \left(\frac{\tau^2 G_{\mathrm{T}} G_{\mathrm{R}} P_{\mathrm{T}}}{(4\pi)^2 P_{\mathrm{c}}} \right)^{\frac{1}{\alpha}} \tag{3}$$

Hence, under the unit disk model, the communication coverage of each node is a perfect disk of radius r. In further discussions, r is not limited to be the transmission range of the unit disk model but a generalized parameter defined by (3).

In effect, the randomness on the received signal power $P_{\mathrm{R}}(d)$ can be described by a function $g(d)$ denoting the probability that a directional communication link exists from transmitter to receiver with distance d. On the basis of $g(d)$, a generic channel model can be defined once $g(d)$ satisfies the following restrictions:

$$\begin{cases} g(d_1) = g(d_2), & \text{if } d_1 = d_2 \quad (4) \\ g(d_1) \le g(d_2), & \text{if } d_1 \ge d_2 \quad (5) \\ 0 < \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(\sqrt{x^2 + y^2}) \mathrm{d}x \mathrm{d}y < \infty \quad (6) \end{cases}$$

The generic channel model has been treated intensively in percolation theory [12,13]. The first restriction indicates that the propagation path is symmetric; the second one indicates that $g(d)$ must be a non-increasing function of d; and the third one avoids the trivial cases that any two nodes are directly connected with probability 1 and that any two nodes are isolated with probability 1. It can be easily shown that both the unit disk model and the log-normal model satisfy these restrictions [14].

As transmit power of each node is actually tunable in many wireless sensor networks [15], we require that during the period of running the proposed distance estimation method, all nodes transmit at a *common* power level, that is, P_{T}. Furthermore, throughout the paper, we make the following assumptions (as is commonly the case in the literature).

Assumption 1. *The attenuations caused by shadowing effects (i.e., Z) between any pairs of nodes are independent and identically distributed.*

Assumption 2. *Communication links are symmetric, namely that node A can directly receive packets from node B as long as node B can directly receive packets from node A.*

Even though field measurements in real applications seem to indicate that the attenuations between two links with a common node are correlated [16], Assumption 1 is generally considered appropriate for far field transmission and is widely used in the literature [14,16–20]. Although the above assumptions may not fully reflect a real network environment, they still enable us to obtain some results as estimates for more realistic situations.

3. THE DISTANCE ESTIMATION METHOD

In this section, we present the method of estimating distances via connectivity and detail its implementation under the log-normal model.

3.1. Estimating distances under the unit disk model

In a static wireless sensor network, given two nodes A and B with coordinates $(x_{\mathrm{A}}, y_{\mathrm{A}})$ and $(x_{\mathrm{B}}, y_{\mathrm{B}})$, their distance is defined to be d ($d \le r$), and two disks with the same radius r represent their individual communication coverage under the unit disk model, as shown in Figure 1. Because of $d \le r$, the two disks intersect and create three disjoint regions. Regarding r as a constant, we define $S = \pi r^2$ and $f(d)$ to be the area of the middle region in Figure 1, where

$$f(d) = \frac{2S}{\pi} \arccos\left(\frac{d}{2r}\right) - d\sqrt{r^2 - \frac{d^2}{4}} \tag{7}$$

It is obvious that the nodes residing in the middle region are common immediate neighbors of A and B, the nodes residing in the left (right) one are non-common immediate neighbors of A (B). Define three random variables M, P, and Q to be the numbers of the three categories of neighbors. Obviously, they are mutually independent

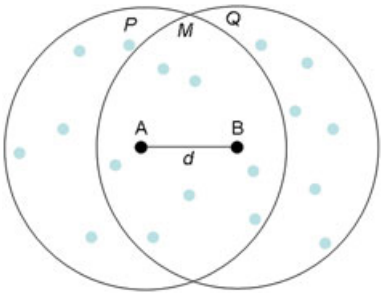


Figure 1. The communication coverage of two nodes under the unit disk model.

and Poisson with means $\lambda f(d), \lambda(S - f(d))$, and $\lambda(S - f(d))$, as pointed out in [13]. The actual values of M, P, and Q can be easily obtained after A and B exchange their neighborhood information. On the basis of the observations of M, P, and Q, an MLE for estimating d is summarized as follows:

Theorem 1. *M, P, and Q are mutually independent Poisson random variables with means $\lambda f(d), \lambda(S - f(d))$, and $\lambda(S - f(d))$, respectively. If $f(d)$ is invertible and S is a non-zero constant, then the MLE for d, termed $\hat{d}$, is*

$$\hat{d} = \begin{cases} f^{-1}(S), & \text{if } M = P = Q = 0; \quad (8) \\ f^{-1}(\hat{\rho}S), & \text{otherwise} \quad (9) \end{cases}$$

where $\hat{\rho} = 2M/(2M + P + Q)$.

Proof. See Appendix A. □

Note that the actual value of λ is not needed in obtaining $\hat{d}$. Although nodes are assumed to follow a random and uniform distribution of density λ (as a result of the Poisson point process), the derivation of the MLE indicates that as long as nodes in a *local* region that covers the communication coverage of two neighboring nodes admits a uniform density, the proposed method is reasonably applicable. Moreover, in some applications, nodes may be placed on the basis of certain regular structures but with noises. For example, in a two-dimensional sensor network, the x-coordinate and y-coordinate of each node are Gaussian with same variance and mean values at one grid point. Compared with a random and uniform distribution, such distribution is even closer to be uniform, and the proposed method can thus attain better performance. In addition, if the least squares method instead of the MLE is applied here, the resulting expression of the distance estimator is actually the same as that in Theorem 1.

3.2. Extension under the generic channel model

Under the generic channel model defined by $g(d)$, M, P, and Q continue to denote the numbers of common and non-common immediate neighbors associated with two nodes. First, we can compute their expectations as follows:

$$E(M + P) = E(M + Q) = \lambda \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \times g(\sqrt{(x - x_B)^2 + (y - y_B)^2}) \mathrm{d}x\mathrm{d}y \quad (10)$$

$$E(M) = \lambda \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(\sqrt{(x - x_A)^2 + (y - y_A)^2}) \times g(\sqrt{(x - x_B)^2 + (y - y_B)^2}) \mathrm{d}x\mathrm{d}y \quad (11)$$

Then, from the third restriction on $g(d)$, that is, (6), it follows that $E(M) < E(M + P) < \infty$. Unlike the unit disk model where the independence among M, P, and Q is straightforward because of having three disjoint regions, the generic channel model does not necessarily lead to such three disjoint regions, while the following theorem guarantees the mutual independence.

Theorem 2. *Suppose a static wireless sensor network is deployed in an infinite plane according to a homogeneous Poisson process of density λ and conforms to the generic channel model defined by $g(d)$; given two nodes in this wireless sensor network, let M be the number of their common immediate neighbors and P and Q be the numbers of their non-common immediate neighbors. Then, M, P, and Q are mutually independent Poisson random variables.*

Proof. See Appendix B. □

Under the generic channel model, S and $f(d)$ are generalized to specify the expectations of M, P and Q, instead of the areas defined under the unit disk model, and have the forms of

$$S = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(\sqrt{(x - x_B)^2 + (y - y_B)^2}) \mathrm{d}x\mathrm{d}y \quad (12)$$

$$f(d) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(\sqrt{(x - x_A)^2 + (y - y_A)^2}) \times g(\sqrt{(x - x_B)^2 + (y - y_B)^2}) \mathrm{d}x\mathrm{d}y \quad (13)$$

Therefore, if S and $f(d)$ satisfy the conditions in Theorem 1, the MLE is directly applicable.

In reality, however, sensor networks are deployed in regions of finite areas, and thus, the expectations of M, P, and Q associated with two nodes, especially those near network boundaries, cannot be derived by simply integrating over an infinite plane to compute S and $f(d)$. This is termed *boundary effects*. In this study, we concentrate on the theoretical foundations of the proposed method and will tackle boundary effects in our future work.

Prior to implementing the proposed method in a static wireless sensor network, it is a premise to know the wireless channel, that is, $g(d)$, such that the quantity S, the function $f(d)$, and its inverse can be determined and then programmed into each node. After that, because of the simple mechanism of the proposed method, a distributed protocol can be easily designed for collecting and exchanging local connectivity information by each node through broadcasting operations. Once a node obtains neighboring information of all its immediate neighbors, this node is able to estimate the distances from its immediate neighbors using the inverse of $f(d)$, S, and the MLE in Theorem 1.

If Assumption 2 holds in a sensor network, each pair of neighboring nodes will have identical information for estimating their distance and thus will obtain the same distance estimate; otherwise, provided that A can hear B but

B cannot hear A, A will estimate their distance, whereas B will not. Such asymmetry in distance estimation can be alleviated by allowing each node to exchange two-hop neighborhood information with its immediate neighbors.

Clearly, if the size of each sensor's neighborhood has the magnitude of $O(1)$, the complexities for communications and computations of running the proposed method in this network are both $O(n)$, where n is the number of nodes in a wireless sensor network, implying that the proposed method is efficient and scalable.

3.3. Implementation under the log-normal model

Provided that the wireless channel is known to be log-normal with known parameters r, σ_{dB}, and α, we shall demonstrate how to determine $g(d)$, S, and $f(d)$ involved in the proposed method.

3.3.1. Formulating $g(d)$

Under the log-normal model, for a transmitter and receiver pair with distance d, if the received signal power described by (1) is no less than P_c, a bi-directional communication link exists between them (as a result of the symmetry in Assumption 2). The probability that the two nodes are able to communicate with each other, that is, $g(d)$, is

$$g(d) = \int_{k \log \frac{d}{r}}^{\infty} \frac{\mathrm{e}^{-\frac{z^2}{2\sigma_{\mathrm{dB}}^2}}}{\sqrt{2\pi}\sigma_{\mathrm{dB}}} \mathrm{d}z \tag{14}$$

where $k = 10\alpha/\log 10$ and log denotes natural logarithm. Evidently, $g(d)$ is determined by r, α, and σ_{dB}, where r can be easily computed given the parameters in (3), and α and σ_{dB} can be derived by measurements obtained prior to the deployment of sensor networks or empirically assigned on the basis of the characteristics of the deployment environment [11]. Alternatively, using the technique presented in [21], α and σ_{dB} can be estimated through processing of RSS measurements (and no distance measurements), and depending on the level of noise and amount of measurement data, they may result in imprecise α and σ_{dB}.

We plot $g(d)$ with respect to d and σ_{dB} given $\alpha = 4$ and $r = 1$ in Figure 2(a). It can be seen that the smaller is d, the higher is the probability that a communication link exists; a larger σ_{dB} tends to inhibit communications for a smaller d but promotes communications for a larger d in comparison with a smaller σ_{dB}.

In view of the restrictions on $g(d)$, it is straightforward to obtain $\lim_{d\to\infty} g(d) = 0$; as such, for an extremely small and positive ε, there exists d_{th} such that $g(d) < \varepsilon$ if $d > d_{\mathrm{th}}$. That is to say, nodes with distances to a node longer than d_{th} hardly communicate with this node directly; as such, d_{th} is a surrogate of the transmission range. This phenomenon can be observed in Figure 2(a).

3.3.2. Formulating S and $f(d)$

In [17,19], the expectations $E(M+P)$ (or $E(M+Q)$) has been well studied. Thus, we can have

$$S = \pi r^2 \mathrm{e}^{\frac{2\sigma_{\mathrm{dB}}^2}{k^2}} \tag{15}$$

By (11), (13), and (14), we can derive the formula for $f(d)$ under the log-normal model. By letting $\alpha = 4$ and $r = 1$, we plot $f(d)$ with respect to different values of d and σ_{dB} in Figure 2(b).

As can be seen in Figure 2(b), $f(d)$ is monotonically decreasing and invertible; hence, Theorem 1 is applicable under the log-normal model. But the closed-form formula for $f(d)$ and its inverse are unavailable. Alternatively, we can establish a piecewise linear function to approximate its inverse; for each affine segment, a linear regression model is applied to predict d. Considering the fact that two nodes with distance longer than d_{th} hardly communicate with each other directly, we restrict the distance estimates to be between 0 and d_{th}. But in a real estimation process, $\hat{\rho}S$ and S may exceed $[f(d_{\mathrm{th}}), f(0)]$ and consequently $\hat{d}$ may exceed $[0, d_{\mathrm{th}}]$. Therefore, we can obtain the distance estimator as follows:

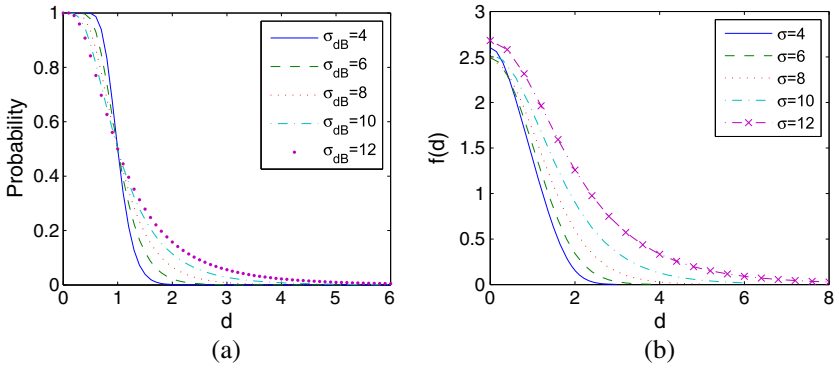


Figure 2. The functions $g(d)$ and $f(d)$ under the log-normal model with $\alpha = 4$ and $r = 1$.

$$\hat{d} = \begin{cases} 0, & \text{if } M = P = Q = 0 \text{ or } \hat{\rho}S > f(0); \\ f^{-1}(\hat{\rho}S), & \text{if } f(d_{\text{th}}) \le \hat{\rho}S \le f(0); \\ d_{\text{th}}, & \text{if } \hat{\rho}S < f(d_{\text{th}}) \end{cases} \tag{16}$$

4. PERFORMANCE ANALYSIS

In this section, we evaluate the performance of the proposed method under the log-normal model from different respects.

4.1. Impact of imprecise α and σ_{dB}

In the proposed method, the parameters of the log-normal model, that is, α and σ_{dB}, are supposed to be known precisely. But what if their values are imprecise? To answer this question, we define

$$\rho_{\alpha,\sigma_{\text{dB}}}(d) = \frac{f(d)}{S} \tag{17}$$

where $f(d)$ and S are computed using (13), (15), and (14) given α and σ_{dB}. Thus, the distance estimator in (16) is equivalently

$$\hat{d} = \begin{cases} 0, & \text{if } M = P = Q = 0 \text{ or } \hat{\rho} > \rho_{\alpha,\sigma_{\text{dB}}}(0); \\ \rho^{-1}_{\alpha,\sigma_{\text{dB}}}(\hat{\rho}), & \text{if } \rho_{\alpha,\sigma_{\text{dB}}}(d_{\text{th}}) \le \hat{\rho} \le \rho_{\alpha,\sigma_{\text{dB}}}(0); \\ d_{\text{th}}, & \text{if } \hat{\rho} < \rho_{\alpha,\sigma_{\text{dB}}}(d_{\text{th}}) \end{cases} \tag{18}$$

Evidently, using imprecise α and/or σ_{dB} results in an incorrect function $\rho^{-1}_{\alpha,\sigma_{\text{dB}}}(\hat{\rho})$, with the result that the accuracy of the distance estimate $\hat{d}$ is degraded. We plot the function $\rho^{-1}_{\alpha,\sigma_{\text{dB}}}(\hat{\rho})$ with respect to different values of α and σ_{dB} in Figure 3.

Supposing σ_{dB} is known to be exactly 4, we investigate the impact of the uncertainty in α. As shown in Figure 3(a), $\rho^{-1}_{\alpha,\sigma_{\text{dB}}}(\hat{\rho})$ is much more sensitive to a small α than to a large α; in other words, for a small α, using an imprecise version of α tends to degrade the accuracy of the distance estimate $\hat{d}$ more seriously than for a large α. Moreover, if α is overestimated, then an underestimated $\hat{d}$ will be produced for a small d but an overestimated $\hat{d}$ for a large d, and vice versa. However, $\rho^{-1}_{\alpha,\sigma_{\text{dB}}}(\hat{\rho})$ does not demonstrate the same sensitivity to σ_{dB} as is observed for α, as illustrated in Figure 3(b). We can conclude that if σ_{dB} is overestimated, then an overestimated $\hat{d}$ will be produced for a small d but an underestimated $\hat{d}$ for a large d, and vice versa.

4.2. Bias and standard deviation

According to Theorem 1, all possible values of $\hat{\rho}$ are rational numbers within $[0, 1]$ so that $\hat{d}$ is a discrete random variable and its jth moment is as follows:

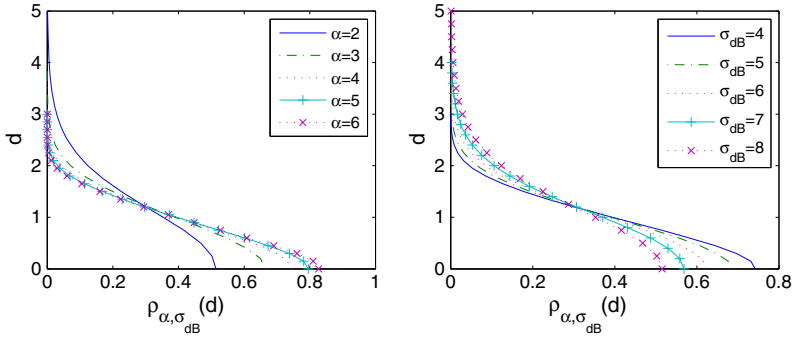
$$E(\hat{d}^j) = \sum_a \left[\hat{d}^j \Pr(\hat{d} = a)\right] \tag{19}$$

We divide the range of $\hat{d}$, that is, $[0, d_{\text{th}}]$, into w equal intervals: $\mathcal{I}_1 = [z_0, z_1), \ldots, \mathcal{I}_w = [z_{w-1}, z_w]$ with $z_i = (i \times d_{\text{th}})/w$. Given a sufficient large w, $\hat{d}$ is approximately constant over each interval, denoted $\tilde{d}_i$. Then we can approximately reformulate (19) as

$$E(\hat{d}^j) \approx \sum_{i=1}^{w} \left[(\tilde{d}_i)^j \Pr(\hat{d} \in \mathcal{I}_i)\right] \tag{20}$$

Toward the probability associated with the ith interval $\mathcal{I}_i$, we have

$$\Pr(\hat{d} \in \mathcal{I}_i) = \begin{cases} \Pr(f(z_1) < \hat{\rho}S \le S)), & \text{if } i = 1; \\ \Pr(f(z_i) < \hat{\rho}S \le f(z_{i-1})), & \text{if } 1 < i < w; \\ \Pr(0 \le \hat{\rho}S \le f(z_{w-1})), & \text{if } i = w \end{cases}$$



(a) $\sigma_{dB} = 4$ (b) $\alpha = 4$

Figure 3. The inverse function of $\rho_{\alpha,\sigma_{\text{dB}}}(d)$ under the log-normal model.

Table I. The values of λ with respect to different σ_{dB} and μ when $\alpha = 4$.

	μ				
σ_{dB}	5	10	20	30	40
0	1.59	3.18	6.37	9.55	12.73
4	1.43	2.86	5.73	8.59	11.45
8	1.04	2.08	4.17	6.25	8.33
12	0.61	1.23	2.45	3.68	4.90

By letting $Y = P + Q$, we have

$$\Pr(b < \hat{\rho}S < c) = \sum_{y=0}^{\infty} [\Pr(b < \hat{\rho}S < c | Y = y) \times \Pr(Y = y)] \tag{21}$$

which makes it possible for us to numerically evaluate the moments of $\hat{d}$ and thus the bias and standard deviation.

Let μ be the expected number of immediate neighbors of a node, namely $\mu = E(M + P) = E(M + Q)$, and the values of λ with respect to different σ_{dB} and μ are listed in Table I. For better presentation, the connectivity index μ will be used in the following discussions instead of the node density λ.

Given $\alpha = 4$, $r = 1$, $w = 1000$, and μ varying from 5 to 40, Figure 4 depicts the numerical bias and standard deviation associated with the proposed method and the corresponding simulation results. The two groups of results are highly consistent, and the comparatively non-smooth aspect of some of the curves, for example, Figure 4(a), is probably attributable to the fact that all observations of M, P, and Q in related results are necessarily integer, and such observations are used in determining the curves.

As shown in Figure 4, the proposed method is obviously *biased*, but the absolute bias is much less than the standard deviation in most cases; except for $\sigma_{\text{dB}} = 0$, the absolute bias and standard deviation are comparable with true distances, especially for short distances and sparse sensor networks, and particularly, when $\mu = 5$, their values are extraordinarily large and nearly twice the corresponding values when $\mu = 10$. Moreover, with μ increasing, the standard deviation always reduces, while the absolute bias reduces in most cases. An intuitive explanation is that with μ increasing, the variances of the ratios $2M/E(2M)$ and $(2M+P+Q)/E(2M+P+Q)$ both decrease, the variance of $\hat{\rho}$ is reduced and so is the variance of $\hat{d}$. As mentioned in the previous section, a large σ_{dB} promotes communications between distant nodes but inhibits communications between close nodes; as a result, connectivity is related to a wide range of distances so that the geometric information implied by connectivity becomes less accurate. Hence, the larger is the σ_{dB}, the worse are both the bias and the standard deviation.

4.3. Root mean square error

As a performance measure, the RMSE is defined to be the square root of the sum of the square bias and variance of estimation errors. We plot the RMSE of $\hat{d}$ produced by the proposed method in Figure 5. As can be seen, the RMSE decreases with μ increasing and σ_{dB} decreasing, which is consistent with how the bias and standard deviation of $\hat{d}$ depend on μ and σ_{dB}. When d is near 0, the RMSE is extraordinarily large compared with the true value of d, implying that the proposed method fails to provide reasonable estimates for short distances. This underperformance with short distances limits the use of the proposed method in practice and is due to a mixed impact of the following facts:

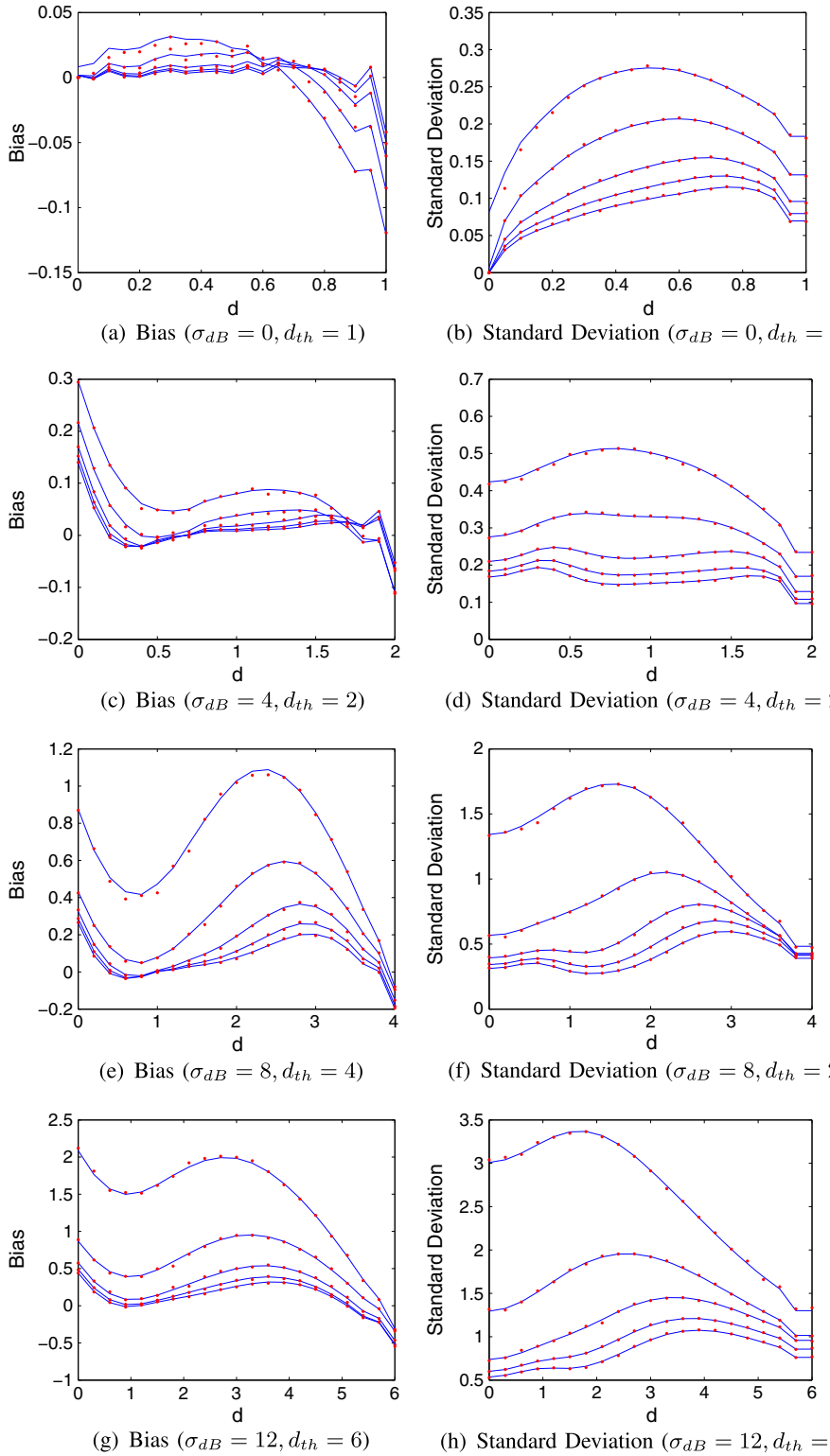
- It is evident that the variance of $2M + P + Q$ is constant no matter what d is, but the variance of $2M$ increases with d decreasing; as a result, $\hat{\rho}$, that is, $2M/(2M + P + Q)$, is more likely to suffer bigger variances when d is small than when d is large.
- As depicted in Figure 3, $\rho^{-1}_{\alpha,\sigma_{\text{dB}}}(\hat{\rho})$ is quite sensitive to $\hat{\rho}$ when d is small, namely that a small perturbation in $\hat{\rho}$ leads to a big change in $\hat{d}$ and thus a big distance estimation error.
- In light of (18), $\hat{d}$ is roughly set 0 when $\hat{\rho}$ is greater than $\rho_{\alpha,\sigma_{\text{dB}}}(0)$, but a small d often causes $\hat{\rho}$ to be within $[\rho_{\alpha,\sigma_{\text{dB}}}(0), 1]$, and hence the underperformance is attained.

To conclude, for short distances, the non-smooth aspect and the sensitivity to $\hat{\rho}$ of the function defined in (18) are responsible for the underperformance.

Under the log-normal model with $\sigma_{\text{dB}} > 0$, distance estimation can be realized by using the RSS measurements, that is, received signal powers. The bias and variance of the resulting distance estimate (denoted $\hat{d}_{\text{RSS}}$) are provided in [22] so that we can compute the RMSE of $\hat{d}_{\text{RSS}}$ and compare it with that of $\hat{d}$ in Figure 5. It can be seen that (i) the RMSE of $\hat{d}_{\text{RSS}}$ increases in direct proportion to d, but that of $\hat{d}$ appears to have comparatively small variations with d increasing, and (ii) the proposed method outperforms the RSS method for long distances by a large margin.

5. ANALYSIS BASED ON THE CRAMÉR–RAO LOWER BOUND

In this section, we formulate the CRLB regarding the distance estimation problem via connectivity, that is, estimating d from M, P, and Q, under the log-normal model. For this estimation problem, the unknown parameters are d and λ. The Fisher Information Matrix (FIM) for this estimation problem, denoted FIM(d, λ), is



(a) Bias ($\sigma_{dB} = 0, d_{th} = 1$) (b) Standard Deviation ($\sigma_{dB} = 0, d_{th} = 1$)

(c) Bias ($\sigma_{dB} = 4, d_{th} = 2$) (d) Standard Deviation ($\sigma_{dB} = 4, d_{th} = 2$)

(e) Bias ($\sigma_{dB} = 8, d_{th} = 4$) (f) Standard Deviation ($\sigma_{dB} = 8, d_{th} = 2$)

(g) Bias ($\sigma_{dB} = 12, d_{th} = 6$) (h) Standard Deviation ($\sigma_{dB} = 12, d_{th} = 6$)

Figure 4. The bias and standard deviation of distance estimation from numerical evaluations (solid lines) and simulations (dashed lines) with $\mu = 5, 10, 20, 30$, and 40, $\alpha = 4$, and $r = 1$. For the standard deviation, a larger μ corresponds to a line to the bottom; for the bias, a larger μ corresponds to a line with the bias closer to 0.

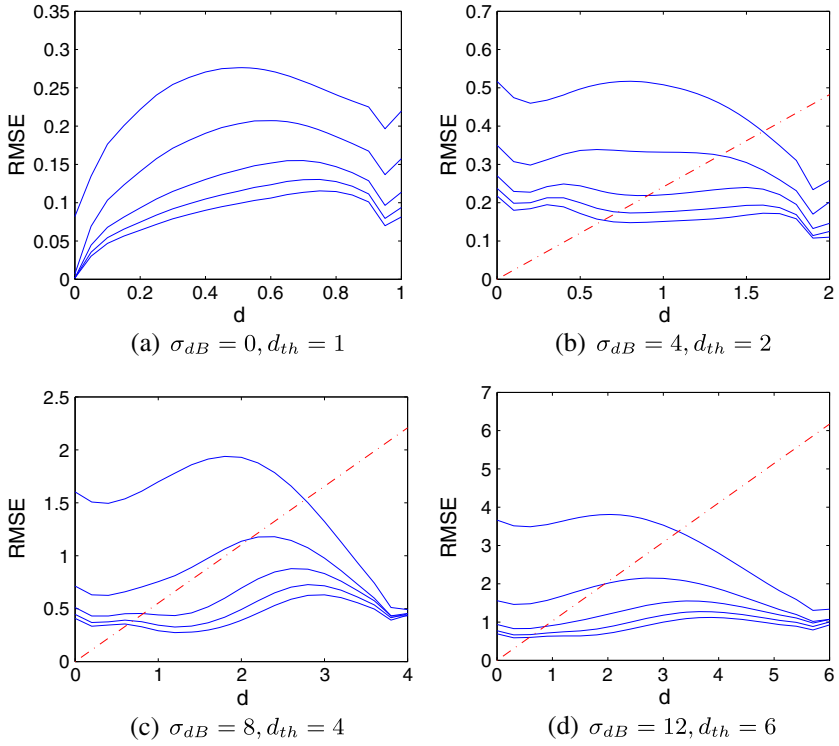


Figure 5. The RMSE of $\hat{d}_{\text{RSS}}$ (dashed lines) and $\hat{d}$ (solid lines; $\mu = 10, 20, 30$, and 40 with $\alpha = 4$ and $r = 1$; a larger μ corresponds to a curve to the bottom) under the log-normal model.

$$\text{FIM}(d, \lambda) = \begin{pmatrix} \lambda \left(f'(d)\right)^2 \left(\frac{1}{f(d)} + \frac{2}{S - f(d)}\right) & -f'(d) \\ -f'(d) & \frac{2S - f(d)}{\lambda} \end{pmatrix} \tag{22}$$

where $f(d)$ is differentiable to the first order in (13) (see Appendix C). Then, the CRLB for d by using any unbiased estimator, denoted CRLB(d), is

$$\text{CRLB}(d) = \frac{(S - f(d))(2S - f(d))f(d)}{2\lambda S^2 (f'(d))^2} \tag{23}$$

Although the CRLB is only valid for unbiased distance estimates and the proposed method is known to be biased, it is still helpful to understand the essential features of the distance estimation problem. In what follows, we shall investigate the influences of various parameters.

5.1. Influence of λ

It is clear that the CRLB is inversely proportional to λ. In other words, better estimation accuracy can be attained in dense wireless sensor networks, which is intuitive and is also illustrated in Figure 5. Hence, it is attractive to apply the proposed method in dense wireless sensor networks. Dense wireless sensor networks, however, are really required in some circumstances. For example, because of the limited energy resource at each node, nodes are usually deployed in high density and may take turns to be active in order to prolong the network lifetime [23]; accordingly, many scheduling strategies have been developed to determine when and which sensors should be powered up and which sensors should be put into energy saving mode while satisfying certain coverage and connectivity requirements [24–29].

5.2. Influence of d

According to (23), it is difficult to directly observe the influence of d on the CRLB, for we do not have the closed-form formulas for $f(d)$ and $f'(d)$ except for the case of $\sigma_{\text{dB}} = 0$. But because it is easily justified that the numerator of (23) is bounded in a narrow range, if the denominator can be very small, the CRLB will be seriously affected by the denominator. On the basis of Figure 2(b), we can obtain some preliminary knowledge about the key component in the denominator, that is, $f'(d)$. As can be seen from Figure 2(b), with d increasing from 0, $|f'(d)|$ firstly experiences a rise and then decreases after d is greater than some value that differs from σ_{dB}; when d increases further, $|f'(d)|$ continuously decreases and approaches 0. Hence, it is postulated that the CRLB will experience a rise with d increasing.

As shown in Figure 4, in the cases of $\sigma_{\text{dB}} = 8$ and 12, the standard deviation displays an evident rise when d is

larger than some value and then drops when d approaches d_{th}. The reason causing such a drop is that we restrict the maximal distance estimate to be d_{th} so that estimates for d near d_{th} are improved. In the case of $\sigma_{\text{dB}} = 4$, $|f'(d)|$ is not so close to 0 when d is near d_{th} but is comparatively small when d is near 0; as a result, the expecting rise does not happen.

5.3. Influences of P_{T} and P_{c}

Provided that G_{T}, G_{R}, and α are known in (3), r is proportional to $(P_{\text{T}}/P_{\text{c}})^{\frac{1}{\alpha}}$, where P_{T} is the transmission power and P_{c} is the threshold of power for communications. If both P_{T} and P_{c} are tunable in wireless sensor networks (which does not break requirement on the common transmission power P_{T}), it will be meaningful to analyze their influences on the CRLB. For simplicity, we shall use r instead of P_{T} and P_{c} to carry out the analysis. To do so, we derive the following theorem.

Theorem 3. *Consider the CRLB that is defined on the basis of (13), (15), and (23), and suppose only d and r are variables. Then, the CRLB for a given distance d with $r = r_0$ is equal to the CRLB for a distance d/r_0 with $r = 1$.*

Proof. See Appendix D. □

This theorem reveals that (i) the CRLB is virtually determined by the ratio d/r and (ii) the value field of the CRLB is invariant no matter how large r is. That is, if the ratio $P_{\text{T}}/P_{\text{c}}$ is raised, distant nodes will tend to become connected so that estimates for long distances will be available, but the CRLB will not exceed the value field of the CRLB associated with the original small value of $P_{\text{T}}/P_{\text{c}}$. Consequently, estimates for long distances will generally suffer less relative errors (i.e., the ratio of the estimation error to d) than those for short distances.

Notice that because the CRLB is not monotonic with d, tuning $P_{\text{T}}/P_{\text{c}}$ does not definitely increase or decrease the corresponding CRLB associated with one given value of d. Because raising $P_{\text{T}}/P_{\text{c}}$ results in more immediate neighbors for each node and consequently more distance estimates, although any distance estimate is not necessarily improved, more available distance estimates will benefit other applications, for example, sensor network localization.

Furthermore, this feature can be exploited in the implementation of the proposed method. Considering the fact that in static wireless sensor networks, the procedure of estimating distances is usually executed only once, and probably in the beginning of the network lifetime, the ratio $P_{\text{T}}/P_{\text{c}}$ can be initially set a high value to achieve a high 'sensor density' by increasing P_{T} and/or decreasing P_{c} and then is tuned to be a normal value after the phase of estimating distances. As a result, more estimates of long distances with comparatively good accuracies will be available.

6. IMPLEMENTING THE METHOD IN PRACTICE

In this section, we improve the proposed method when dealing with short distances and then test it in a practical environment.

6.1. Dealing with short distances

Given α and σ_{dB}, define $\epsilon_{\alpha,\sigma_{\text{dB}}}$ to be the RMSE when $d = 0$. As illustrated in Figure 5, the RMSE of distance estimates produced by the proposed method experiences small variations as d increases from 0 up to d_{th}, so that if $d \geq \epsilon_{\alpha,\sigma_{\text{dB}}}$ the RMSE tends to be under d, implying that relatively good performance is attained. Moreover, on the grounds of the analysis in Section 4.3, we focus on the function defined by (18) with $\hat{\rho} \in [\rho_{\alpha,\sigma_{\text{dB}}}(\epsilon_{\alpha,\sigma_{\text{dB}}}), 1]$ and reformulate it by a linear function $[(1-\hat{\rho})\epsilon_{\alpha,\sigma_{\text{dB}}}]/[1-\rho_{\alpha,\sigma_{\text{dB}}}(\epsilon_{\alpha,\sigma_{\text{dB}}})]$ that smoothly transforms any $\hat{\rho}$ between $\rho_{\alpha,\sigma_{\text{dB}}\text{dB}}(\epsilon_{\alpha,\sigma_{\text{dB}}})$ and 1 to a distance estimate between 0 and $\epsilon_{\alpha,\sigma_{\text{dB}}}$. Consequently, (18) is updated to be

$$\hat{d} = \begin{cases} 0, & \text{if } M = P = Q = 0; \\ \frac{(1-\hat{\rho})\epsilon_{\alpha,\sigma_{\text{dB}}}}{1-\rho_{\alpha,\sigma_{\text{dB}}}(\epsilon_{\alpha,\sigma_{\text{dB}}})}, & \text{if } \hat{\rho} > \rho_{\alpha,\sigma_{\text{dB}}}(\epsilon_{\alpha,\sigma_{\text{dB}}}); \\ \rho^{-1}_{\alpha,\sigma_{\text{dB}}}(\hat{\rho}), & \text{if } \rho_{\alpha,\sigma_{\text{dB}}}(d_{\text{th}}) \leq \hat{\rho} \\ & \quad \leq \rho_{\alpha,\sigma_{\text{dB}}}(\epsilon_{\alpha,\sigma_{\text{dB}}}); \\ d_{\text{th}}, & \text{if } \hat{\rho} < \rho_{\alpha,\sigma_{\text{dB}}}(d_{\text{th}}) \end{cases} \tag{24}$$

6.2. Test the method in a real environment

In [30], a sensor network consisting of 44 nodes was deployed in a real environment, and RSS measurements between any two nodes were reported. On the basis of their measurement data, we can simulate a realistic environment to implement our method. According to [30], $\alpha = 2.3$, $\sigma_{\text{dB}} = 3.92$, and $\overline{P_{\text{R}}}(R_0) = -37.47$ dBm. But to proceed with the experiment, we also need to specify the threshold power P_{c}, which essentially defines whether two nodes are 'connected', and $\epsilon_{\alpha,\sigma_{\text{dB}}}$ in (24). After that, we can compute the function $g(d)$ associated with this channel and then obtain the distance estimators on the basis of (18) and (24), respectively. To avoid boundary effects as much as possible, we consider the four nodes near the center of the deployment region, that is, nodes 15, 23, 24, and 25, and only estimate the inter-node distances associated with the four nodes by using the originally proposed method and the method with the adjustment.

In this experiment, by letting $\epsilon_{\alpha,\sigma_{\text{dB}}}$ be $0.5r$ and raising P_{c} from -61 to -52 dBm, the average distance estimation errors incurred by the original and adjusted methods are listed in Table II. By the distance estimates produced by the RSS method (which were also provided by [30]), we compute the corresponding average distance estimation error, that is, 1.07 m. As depicted in the table, (i) the adjusted method always outperforms the original method; (ii) the

Table II. The experimental results with respect to different P_c.

P_c (dB m)	r (m)	ND	DEE (m)	DEE adjusted (m)
−52	4.28	8.50	1.00	0.44
−53	4.73	11.25	1.49	0.56
−54	5.23	14.00	1.95	0.95
−55	5.78	17.25	1.78	0.74
−56	6.38	20.75	1.81	0.70
−57	7.05	27.25	2.00	1.08
−58	7.80	31.75	2.08	1.41
−59	8.62	34.50	2.11	1.57
−60	9.52	37.00	2.15	1.72
−61	10.53	38.75	2.14	1.67

ND, average node degree; DEE, average distance estimation error.

original method outperforms the RSS method only in the case of $P_c = -52$ dBm, whereas the adjusted method outperforms the RSS method if P_c (dBm) is between −56 and −52 dBm; and (iii) although the average node degree increases with P_c decreasing, the average error obtained with the adjusted method increases in general, a phenomenon which is attributable to boundary effects.

7. APPLYING THE METHOD IN LOCALIZATION

In this section, we report the improvement in connectivity-based sensor localization by using the proposed method.

7.1. Connectivity-based localization algorithms

Connectivity-based localization algorithms, for example, [6,8], generally involve as a crucial component a mechanism of converting connectivity information into rough distance estimates, which are then used for localization as in distance-based localization algorithms.

The DV-hop scheme proposed in [6] employs distance vector exchange. Both sensor and anchor exchange distance tables that contain the locations of and the hop counts to anchors with their corresponding neighboring nodes. Once an anchor obtains the distance tables from other anchors, it estimates an individual average distance per hop and broadcasts this average distance into the network. A sensor approximates its geographic distance to an anchor by multiplying the hop count to this anchor and the associated average distance per hop and then estimates its location by performing trilateration if a sufficient number of distance estimates are obtained. Its variant, that is, the DV-distance scheme, is almost the same as the DV-hop scheme except that it employs the geographic distances measured with the use of radio signals other than hop counts.

Multi-dimensional scaling map (MDS-MAP) proposed in [8] approximates the distance between two nodes by the length of their shortest path and then uses multidimensional scaling to generate a relative map that represents the relative positions of nodes. Once a sufficient number of anchors are known, MDS-MAP estimates the absolute coordinates of all the sensors in the network. Like DV-distance, MDS-MAP can also employ geographic distance measurements; we term it MDS-MAP distance.

In both DV-hop and MDS-MAP, the distance between two nodes is roughly estimated according to the length of the shortest path between them, namely that the one-hop distances along any shortest path are assumed to be equal. As opposed to this assumption, our method provides comparatively accurate estimates of one-hop distances and thus helps to improve the quality of connectivity-based sensor localization, which will be demonstrated in the following subsection.

Many methods have been developed in the literature to improve DV-hop. For instance, in [31], estimating the distance from a sensor to an anchor not only uses the length of the shortest path between them as in DV-hop but also exploits the lengths of the shortest paths from this sensor's immediate neighbors to the anchor. Moreover, DV-hop suffers large errors in anisotropic networks, because the estimates of distances from sensors to anchors can be extraordinarily inaccurate. Accordingly, [32–34] were developed to alleviate the impacts of the anisotropic network topology on the estimates of distances. Because comparatively accurate estimates of one-hop distances provided by our method are the basis for estimating distances from sensors to anchors, it is attractive to combine our method with these DV-hop related methods to improve the estimates of distances from sensors to anchors and thus to improve localization accuracy.

7.2. Simulations

We simulate connectivity-based sensor localization under the log-normal model using DV-hop and MDS-MAP and their distance-based counterparts DV-distance and MDS-MAP distance (with distance measurements from the adjusted method).

To avoid boundary effects, we actually generate wireless sensor networks over a large square with side of 18 but only localize the nodes inside of a small one with side of 6 and concentric to the large one. However, the nodes outside of the small one are sometimes used in estimating distances between nodes inside. Four nodes inside the small square and closest to its four corners are chosen as anchors. Regarding the constants parameterizing the log-normal model, α is known to be 4, σ_{dB} takes values from $0, 4, 8, 12$, and P_T and P_c are properly assigned such that $r = 1$. Furthermore, λ takes proper values such that μ varies from 10 to 40 with step size 5 (see Table I).

For each choice of σ_{dB} and μ, 100 independent runs are carried out. In each run, first, a static wireless sensor network is generated according to a homogeneous Poisson process of density λ; second, distance between any pair of neighboring nodes is estimated on the basis of Theorem 1, and accordingly, the average absolute distance estimation

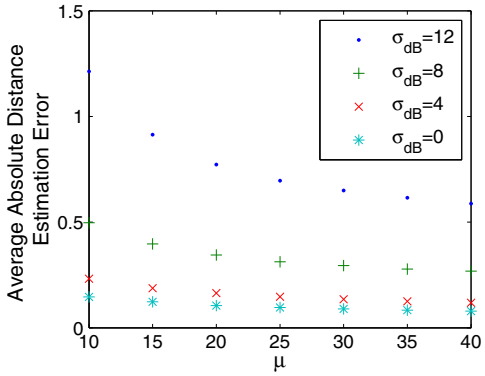


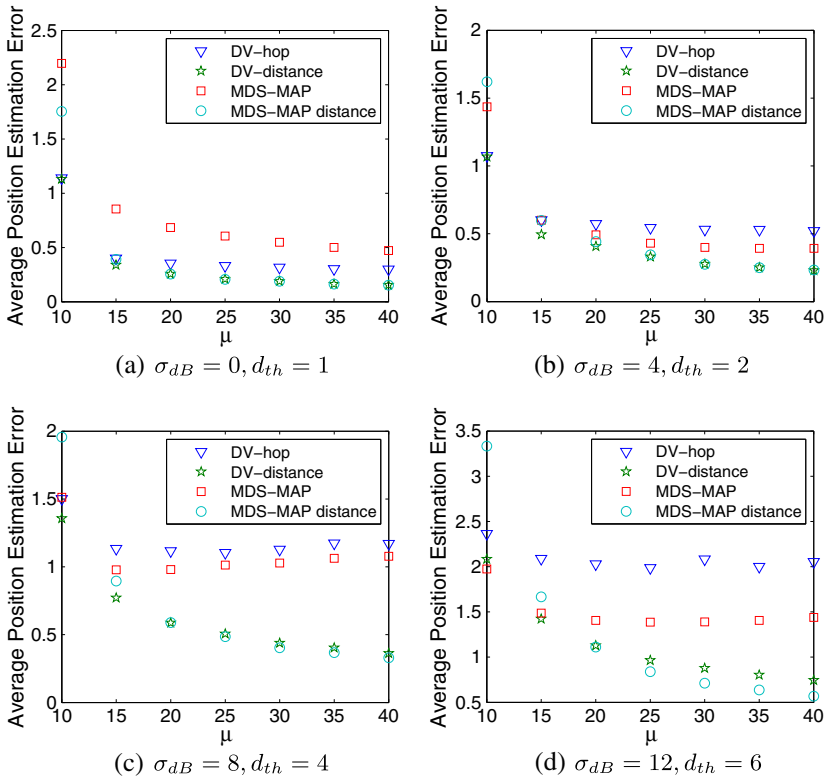
Figure 6. Average absolute distance estimation errors with $\alpha = 4$ and $r = 1$.

error is computed; third, sensors are localized by using four localization algorithms: DV-hop, MDS-MAP, and their distance-based counterparts DV-distance and MDS-MAP distance (with distances coming from the second step); and finally, the average absolute distance estimation error and the average position estimation error are computed for each localization algorithm. Then, the average absolute distance estimation errors and the average position estimation errors are averaged over the 100 independent runs and plotted in Figures 6 and 7.

DV-hop and DV-distance are almost the same except that DV-hop assumes identical one-hop distances along any shortest path, whereas DV-distance uses distance estimates with comparatively good accuracies; this is also true for MDS-MAP and MDS-MAP distance. Because only connectivity information with the assistance of four anchors is exploited to realize sensor localization in the simulations, the fact that DV-distance and MDS-MAP distance use distance estimates produced by our method and their superior performance imply the advantages of our method.

8. CONCLUSIONS

In this paper, we proposed the method of estimating distances via connectivity in static wireless sensor networks by dealing with a generic channel model, including the realistic log-normal model. The proposed method is not relying on extra hardware, totally distributed and energy efficient due to its simple mechanism and computations. Under the log-normal model, the bias and standard deviation of distance estimates from the proposed method were numerically evaluated and verified by simulations; the proposed method outperforms the RSS method for long distances; a CRLB analysis was carried out for the problem of estimating distances using connectivity, and useful guidelines for implementing the proposed method were derived; the influences of uncertainties in the log-normal



(a) $\sigma_{dB} = 0, d_{th} = 1$ (b) $\sigma_{dB} = 4, d_{th} = 2$

(c) $\sigma_{dB} = 8, d_{th} = 4$ (d) $\sigma_{dB} = 12, d_{th} = 6$

Figure 7. Average position estimation errors with $\alpha = 4$ and $r = 1$.

model were examined; and finally, extensive simulations confirmed the advantages of the proposed method. In future work, we may tackle estimating distances involving nodes near the network boundaries, implement the method in a realistic environment and combine the method with the DV-hop related methods to make further improvement.

APPENDIX A: PROOF OF THEOREM 1

Establish a statistical model: observations of M, P and Q provide measured data $\phi = [\,m \quad p \quad q\,]$ where m, p, q are non-negative integers; the unknown parameters are $\theta = [\,d \quad \lambda\,]$. By formulating the likelihood function, we can derive that the MLE for d is the solution for d in the following equation set:

$$\begin{cases} \dfrac{m}{f(d)} - \dfrac{p+q}{S-f(d)} + \lambda = 0, & (25) \\ \dfrac{m+p+q}{\lambda} - (2S - f(d)) = 0 & (26) \end{cases}$$

By eliminating λ, we can obtain

$$2mS = (2m + p + q) f(d) \tag{27}$$

If $2m + p + q > 0$, $\hat{d} = f^{-1}\{[2m/(2m + p + q)]S\}$; otherwise, the solution for d is not well defined, but because $f(d) = S$ maximizes the likelihood, we have $\hat{d} = f^{-1}(S)$. Thus, we prove the theorem.

APPENDIX B: PROOF OF THEOREM 2

Let the two nodes be A and B and consider a finite region of area D covering A and B. As pointed out in [13], apart from A and B, the rest of the Poisson process is not affected, namely that the number of remaining nodes in this region, denoted N, is still Poisson with mean λD. Choosing an arbitrary node C from the N nodes, one of the following cases hold: (i) C can directly communicate with A and B; (ii) C can directly communicate with A but not B; (iii) C can directly communicate with B but not A; and (iv) C cannot directly communicate with A or B.

Conduct a trial for each of the N nodes to decide how it communicates with A and B, and each trial results in the above four cases with probabilities p_1, p_2, p_3, and p_4, respectively. Because of the independence of connectivity assumed in Assumption 1, the N trials are then independent from each other. Evidently, M, P, and Q represent the numbers of nodes belonging to the first three cases. In addition, let the random variable L denote the number of nodes belonging to the last case. Because of $p_1 + p_2 + p_3 + p_4 = 1$, M, P, Q, and L follow a multinomial distribution with parameters N and p_1, p_2, p_3, and p_4. Considering N is Poisson with mean λD, from the theorem on page 8 in [13], it follows that M, P, Q, and L are mutually independent Poisson random variables with means $\lambda D p_1, \lambda D p_2, \lambda D p_3$, and $\lambda D p_4$, respectively. Now, we let the region approach the infinite plane and can conclude that M, P, and Q are mutually independent Poisson with finite means.

APPENDIX C: THE EXISTENCE OF $f'(d)$

At first, consider the following expression:

$$\begin{aligned} &\lim_{\varepsilon\to 0}\left(-\frac{1}{\varepsilon}\left[g(\sqrt{x^2+d^2-2xd\cos\theta}) \right.\right. \\ &\qquad \left.\left. -g(\sqrt{x^2+(d+\varepsilon)^2-2x(d+\varepsilon)\cos\theta})\right]\right) \\ &= \frac{k(x\cos\theta - d)\mathrm{e}^{\frac{-(k(\log\sqrt{x^2+d^2-2xd\cos\theta}-\log r))^2}{2\sigma_{\mathrm{dB}}^2}}}{\sqrt{2\pi}\sigma_{\mathrm{dB}}(x^2+d^2-2xd\cos\theta)} \end{aligned}$$

which is bounded for $x \in [0, +\infty)$. Moreover, the derivative of $f(d)$ can be formulated as

$$\begin{aligned} f'(d) = \int_0^\infty \int_0^{2\pi} &\times g(x)x \lim_{\varepsilon\to 0}\left(-\frac{1}{\varepsilon}\left[g(\sqrt{x^2+d^2-2xd\cos\theta}) \right.\right. \\ &\left.\left. -g(\sqrt{x^2+(d+\varepsilon)^2-2x(d+\varepsilon)\cos\theta})\right]\right)\mathrm{d}\theta\mathrm{d}x \end{aligned} \tag{28}$$

Because $\int_0^\infty \int_0^{2\pi} g(x)x\mathrm{d}\theta\mathrm{d}x$ equals $E(M + P)$ and is convergent, $f'(d)$ is also convergent.

APPENDIX D: PROOF OF THEOREM 3

Regarding r as a variable, we substitute the notations as follows: $S \to S(r)$, $f(d) \to f(r, d)$, $f'(d) \to \partial f(r, d)/\partial d$, $\mathrm{CRLB}(d) \to \mathrm{CRLB}(r, d)$, $g(d) \to g(r, d)$. According to (14), we have

$$g(r, dr) = \int_{k\log d}^{\infty} \frac{\mathrm{e}^{-\frac{z^2}{2\sigma_{\mathrm{dB}}^2}}}{\sqrt{2\pi}\sigma_{\mathrm{dB}}}\mathrm{d}z = g(1, d)$$

By (13), we have

$$\begin{aligned} f(r,dr) &= \int_0^\infty \int_0^{2\pi} g(r,x) \\ &\quad \times g(r,\sqrt{x^2+(dr)^2-2xdr\cos\theta})x\mathrm{d}\theta\mathrm{d}x \\ &= \int_0^\infty \int_0^{2\pi} g(r,xr) \\ &\quad \times g(r,r\sqrt{x^2+d^2-2xd\cos\theta})r^2x\mathrm{d}\theta\mathrm{d}x \\ &= r^2\int_0^\infty \int_0^{2\pi} g(1,x) \\ &\quad \times g(1,\sqrt{x^2+d^2-2xd\cos\theta})x\mathrm{d}\theta\mathrm{d}x \\ &= r^2 f(1,d) \end{aligned}$$

Moreover, we can obtain

$$\left.\frac{\partial f(x,y)}{\partial y}\right|_{x=1,y=d} = \frac{1}{r}\times \left.\frac{\partial f(x,y)}{\partial y}\right|_{x=r,y=dr} \quad (29)$$

By $S(r) = r^2 S(1)$ (based on (15)), (23) and the above formulas, we can obtain

$$\begin{aligned} &\mathrm{CRLB}(r,dr) = \mathrm{CRLB}(1,d) \text{ equivalently,} \\ &\mathrm{CRLB}(r,d) = \mathrm{CRLB}\left(1,\frac{d}{r}\right) \end{aligned} \quad (30)$$

ACKNOWLEDGEMENTS

The work was supported by ARC (Australian Research Council) under DP-110100538 and NICTA. C. Yu is an ARC Queen Elizabeth II Fellow and is also supported by the Overseas Expert Program of Shandong Province. B. Huang was with the Australian National University when the work was performed, and he was also supported by ARC, NICTA and the Key Laboratory of Computer Networks of Shandong Province. NICTA is funded by the Australian Government as represented by the Department of Broadband, Communications and the Digital Economy and the Australian Research Council through the ICT Centre of Excellence program. This material is based on research sponsored by the Air Force Research Laboratory, under agreement number FA2386-09-1-4136. The US Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the Air Force Research Laboratory or the US Government.

REFERENCES

1. Mao G, Fidan B, Anderson BDO. Wireless sensor network localization techniques. *Computer Networks* 2007; **51**(10): 2529–2553.
2. Li L, Halpern JY, Bahl P, Wang Y, Wattenhofer R. A cone-based distributed topology-control algorithm for wireless multi-hop networks. *IEEE/ACM Transactions on Networking* 2002; **13**: 147–159.
3. Siripongwutikorn P, Thipakorn B. Mobility-aware topology control in mobile ad hoc networks. *Computer Communications* 2008; **31**(14): 3521–3532.
4. Sitanayah L, Datta A, Cardell-Oliver R. Heuristic algorithm for finding boundary cycles in location-free low density wireless sensor networks. *Computer Networks* 2010; **54**(10): 1630–1645.
5. Bulusu N, Heidemann J, Estrin D. GPS-less low cost outdoor localization for very small devices. *IEEE Personal Communications Magazine* 2000; **7**(5): 28–34.
6. Niculescu D, Nath B. Ad hoc positioning system (APS), In *Proc. of the IEEE Globecom*, San Antonio, TX, USA, 2001; volume 5, 2926–2931.
7. Nagpal R, Shrobe H, Bachrach J. Organizing a global coordinate system from local information on an ad hoc sensor network, In *Proc. of the ACM/IEEE IPSN*, Palo Alto, CA, USA, 2003; 333–348.
8. Shang Y, Ruml W, Zhang Y, Fromherz MPJ. Localization from mere connectivity, In *Proc. of the ACM MobiHoc*, Annapolis, Maryland, USA, 2003; 201–212.
9. Buschmann C, Pfisterer D, Fischer S. Estimating distances using neighborhood intersection, In *Proc. of the IEEE Emerging Technologies and Factory Automation*, Prague, Czech, 2006; 314–321.
10. Buschmann C, Hellbrück H, Fischer S, Kröller A, Fekete S. Radio propagation-aware distance estimation based on neighborhood comparison, In *Proc. of the 4th European Conference on Wireless Sensor Networks*, Delft, The Netherlands, 2007; 325–340.
11. Rappaport T. *Wireless Communications: Principles and Practice*. Prentice Hall PTR: New Jersey, USA, 2001.
12. Meester R, Roy R. *Continuum Percolation*. Cambridge University Press: Cambridge, UK, 1996.
13. Franceschetti M, Meester R. *Random Networks for Communication: From Statistical Physics to Information Systems*. Cambridge University Press, 2007.
14. Ta X, Mao G, Anderson BDO. On the Connectivity of Wireless Multi-hop Networks with Arbitrary Wireless Channel Models. *IEEE Communications Letters* 2009; **13**(3): 181–183.

15. Ramaswami R, Parhi KK. Distributed scheduling of broadcasts in a radio network, In *Proc. of the IEEE INFOCOM*, Ottawa, Ont., Canada, 1989; 497–504 vol.2.
16. Mukherjee S, Avidor D. Connectivity and transmit-energy considerations between any pair of nodes in a wireless ad hoc network subject to fading. *IEEE Transactions on Vehicular Technology* 2008; **57**(2): 1226–1242.
17. Orriss J, Barton SK. Probability distributions for the number of radio transceivers which can communicate with one another. *IEEE Transactions on Communications* 2003; **51**(4): 676–681.
18. Bettstetter C, Hartmann C. Connectivity of wireless multihop networks in a shadow fading environment. *Wireless Networks* 2005; **11**(5): 571–579.
19. Miorandi D, Altman E. Coverage and connectivity of ad hoc networks in presence of channel randomness, In *Proc. of the IEEE INFOCOM*, Miami, FL, USA, 2005; 491–502.
20. Miorandi D. The impact of channel randomness on coverage and connectivity of ad hoc and sensor networks. *IEEE Transactions on Wireless Communications* 2008; **7**(3): 1062–1072.
21. Mao G, Anderson BDO, Fidan B. Path loss exponent estimation for wireless sensor network localization. *Comput. Netw.* 2007; **51**: 2467–2483.
22. Chitte SD, Dasgupta S, Ding Z. Distance estimation from received signal strength under log-normal shadowing: bias and variance. *IEEE Signal Processing Letters* 2009; **16**(3): 216–218.
23. Akyildiz IF, Su W, Sankarasubramaniam Y, Cyirci E. Wireless sensor networks: a survey. *Computer Networks* 2002; **38**(4): 393–422.
24. Pan J, Hou YT, Cai L, Shi Y, Shen SX. Topology control for wireless sensor networks, In *Proc. of the ACM MobiCom*, San Diego, CA, USA, 2003; 286–299.
25. Chen B, Jamieson K, Balakrishnan H, Morris R. Span: an energy-efficient coordination algorithm for topology maintenance in ad hoc wireless networks. *Wireless Networks* 2002; **8**(5): 481–494.
26. Wang X, Xing G, Zhang Y, Lu C, Pless R, Gill C. Integrated coverage and connectivity configuration in wireless sensor networks, In *Proc. of the ACM SenSys*, Los Angeles, CA, USA, 2003; 28–39.
27. Shang Y, Shi H. A new density control algorithm for WSNs, In *Proc. of the IEEE LCN*, Tampa, FL, USA, 2004; 577–578.
28. Ye F, Zhong G, Cheng J, Lu S, Zhang L. PEAS: A robust energy conserving protocol for long-lived sensor networks, In *Proc. of the IEEE ICDCS*, Providence, RI, USA, 2003; 28–37.
29. Yen L, Cheng Y. Range-based sleep scheduling (RBSS) for wireless sensor networks. *Wireless Personal Communications* 2009; **48**(3): 411–423.
30. Wireless sensor network localization measurement repository. http://www.eecs.umich.edu/hero/localize/, 2011.
31. Ma D, Meng JE, Wang B, Lim HB. A novel approach towards source-to-destination distance estimation in wireless sensor networks, In *Proc. of the Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP)*, Melbourne, Australia, 2009; 463–467.
32. Lederer S, Wang Y, Gao J. Connectivity-based localization of large scale sensor networks with complex shape, In *Proc. of the IEEE INFOCOM*, Phoenix, AZ, USA, 2008; volume 5, 789–797.
33. Lim H, Hou JC. Distributed localization for anisotropic sensor networks. *ACM Transactions on Sensor Networks* 2009; **5**: 1–26.
34. Wang Y, Li K, Wu J. Distance estimation by constructing the virtual ruler in anisotropic sensor networks, In *Proc. of the IEEE INFOCOM*, San Diego, CA, USA, 2010; 1172–1180.

AUTHORS' BIOGRAPHIES

Baoqi Huang received his BEng degree from Inner Mongolia University, China, and his MS degree from Peking University, China, both in computer science. Between March 2008 and September 2011, he was pursuing his PhD degree in information engineering from the Australian National University, Canberra, Australia. Currently, he is with the school of computer, Inner Mongolia University, China. His research interests include wireless sensor networks and mobile *ad hoc* networks.

Changbin (Brad) Yu received his BEng degree with first class honors in computer engineering from Nanyang Technological University, Singapore, in 2004 and his PhD in information engineering from the Australian National University, Canberra, Australia, in 2008. He is now a an ARC Queen Elizabeth II Fellow at the Australian National University and adjunct at NICTA Ltd and Shandong Computer Science Center. He was a recipient of an ARC Australian Postdoctoral Fellowship in 2008, the Chinese Government Award for Outstanding Chinese Students Abroad in 2006, the Australian Government's Endeavour

Asia Award in 2005, and an ARC Queen Elizabeth II Fellowship in 2011. His current research interests include control of autonomous formations, multi-agent systems, sensor networks, and graph theory.

Brian D. O. Anderson is a distinguished professor at the Australian National University and a distinguished researcher in National ICT Australia. He obtained his PhD in electrical engineering from Stanford University in 1966. He is a fellow of the IEEE, IFAC, the Australian Academy of Science, the Australian Academy of Technological Sciences and Engineering, and the Royal Society and a foreign associate of the National Academy of Engineering. His current research interests are in distributed control, including control of UAV formations, sensor networks, and econometric modeling.

Guoqiang Mao received his bachelor's degree in electrical engineering, his master's degree in engineering, and his PhD degree in telecommunications engineering in 1995, 1998, and 2002, respectively. In December 2002, he joined the School of Electrical and Information Engineering, the University of Sydney, Australia, where he is a senior lecturer now. He has published more than 50 papers in prestigious journals and refereed conference proceedings. He was listed in the 25th Anniversary Edition of Marquis 'Who's Who in the World' (2008) and in the 9th (2007) and 10th (2008) Anniversary Editions of Marquis 'Who's Who in Science and Engineering' for 'exceptional achievements in Science and Engineering'. His research interests include wireless localization techniques, wireless multihop networks, graph theory, and its application in networking, telecommunications traffic measurement, analysis and modeling, and network performance analysis.